1965

Optimum finite memory nonlinear filter

Ching-Chyoun Yang

OPTIMUM FINITE MEMORY NONLINEAR FILTER

BY

CHING-CHYOUN YANG, 1940

A

THESIS

Submitted to the faculty of the

THE UNIVERSITY OF MISSOURI AT ROLLA

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

Rolla, Missouri

1965

Approved

by

J R Betten (Advisor) S J Pagano

G G Skitek L L Walters

ABSTRACT

This thesis investigates the possibility of using the Zadah theory and a machine type computer to design an optimum finite memory nonlinear filter for extraction of Poisson distributed signal pulses from a background of Guassian noise. Basically the procedure involves the technique of solving simultaneous equations by the computer.

We demonstrate that it is possible to find the limitation of the filter of class η_1 in terms of an absolute minimum mean square error. However, due to the capacity of the computer we do not actually determine the absolute minimum. Nevertheless we do show that the mean square error decreases with increasing memory time T.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. J. R. Betten, Professor of Electrical Engineering, for providing valuable guidance and suggestions during the course of this investigation.

The author also wishes to thank Professor G. G. Skitek and Professor S. J. Pagano for their valuable suggestions.

The author would also like to thank Mrs. Roberta Rahm for her typing the whole thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. Background

The problem of designing an optimum linear filter and predictor was first conceived by Kolmogoroff in 1942 and shortly afterwards was solved by Wiener (1). Though Wiener's interest almost immediately turned to nonlinear processing (2) and though he devoted much brilliant work to it (3), the problem of an optimum nonlinear filter did not prove mathematically tractable. Wiener and his collaborators Y. W. Lee (4) and A. G. Bose (5) as well as Singleton (6) have made significant contributions to it.

Another principal investigator in the area of nonlinear filter is Zadah (7) and (8). Zadah's work is quite different from Wiener, et al. Wiener dealt essentially with a series of functionals as a type of representation while Zadah dealt with a single functional as a type of representation. More specifically Zadah considers the theory of optimum nonlinears in terms of classes η_1, η_2, η_3 - - -, of filters such that each class in the sequence includes all the preceding classes and furthermore, the class of linear filters is a subclass of every class in the sequence. A filter of class η_m is described in terms of a characteristic functional which involves m age variables and m values of of the input time function. For the class η_m, the filter output y(t) is expressed as

$$y(t) = \int_0^{\infty}\int_0^{\infty}---\int_0^{\infty} K\left[x(t-\tau_1),\ x(t-\tau_2),\ ---,\ x(t-\tau_m),\right.$$

$$\left.\tau_1, \tau_2, \tau_3 - - - \tau_m\right] d\tau_1 d\tau_2 d\tau_3 --- d\tau_m$$

where: x(t) is the input to the filter and the τ's are dummy variables of integration. The function $K[x(t-\tau_1), x(t-\tau_2), --- x(t-\tau_m), \tau_1, \tau_2 --- \tau_m]$ is any real function of its arguments and is called the characteristic function of the filter with which it is associated.

The nonlinear filter is simply defined as a filter which is not linear. For this reason, it would be expecting too much if one hoped to find a universal procedure of designing an optimum nonlinear filter for all different kinds of statistical inputs. However, as defined by Zadah, the nonlinear filter can be thought of in successive classes $\eta_1, \eta_2, - - - \eta_m$. The optimum filter within a class η_m is in general superior to and more complex than the optimum filters within the class $\eta_1, \eta_2, \eta_3 - - - \eta_{m-1}$, and the class of nonlinear filters. The amount of statistical data needed for determining the optimum filter within a class η_m increases with m.

There are many ways of defining an optimum filter. The most common one is the least mean square error criterion used by Wiener. The determination of the minimized mean square error within a class η_m requires the knowledge of 2mth order probability distribution function for the signal and noise.

The theory outlined in Zadah's work does not usually lead to explicit expressions for the characteristic functions of optimum filters within the classes η_1, η_2, η_3, - - -. It leads instead to a sequence of linear integral equations which are satisfied by the respective characteristic functions. Thus, the determination of the optimum filter within a specified class η_m is reduced to the solution of a linear integral equation of the first kind and 2m order. Except for m=1, this system of equations requires machine type solutions.

The simplest kind of nonlinear filter of Zadah's type is the so-called zero-memory (where m=1) nonlinear filter. It is described as $y(t) = f[x(t)]$, where $y(t)$ and $x(t)$ denote the filter output and input respectively, and "f" is the transfer characteristic.

Wang (9) applied the results of Zadah's work to a specific problem in which a nonlinear filter was required because a linear filter was essentially ineffective in separating signal from noise. The ineffectiveness stemmed from the fact that the autocorrelation function of the signal plus noise had essentially the same variational character as the signal alone. Therefore, the transfer function of the linear filter turned out to be simply a constant. Thus, the output of the linear filter was just an attenuated version of the input and no virtual separation of signal from noise was obtained. However, in order to avoid the machine type computation, Wang confined his attention to a zero-memory nonlinear filter.

He used the desired output as the signal itself and obtained the following characteristic function for the optimal zero-memory nonlinear filter:

$$f(x) = \frac{\int_{-\infty}^{\infty} s \cdot P(x,s)\,ds}{P(x)}$$

From this equation it is clear that the only information needed in the optimization of a zero-memory nonlinear filter is the joint second order probability density function of the input and desired output. Wang also shows that the optimum nonlinear zero-memory filter provides a finite mean square error which represents a definite improvement over the optimum linear filter.

B. Purpose

The purpose of this thesis is to extend the work of Wang (op. cit.) to yield a filter which is capable of doing a better job (in the minimum mean square error sense) of separating signal from noise than the zero memory filter designed by Wang.

Although the filter to be considered here is in the same class η_1 as the zero memory filter, it is of greater capability because it possesses a finite non-zero memory. By memory is meant the capability of operating upon a finite portion (duration) of the input signal in order to produce one output at each instant of time. Generally speaking, the larger the filter memory the smaller will be the corresponding mean square error. The relation between the memory time, T, and the mean square error $\overline{e^2(t)}$ will also be investigated here.

It is important to point out, however, that the solutions presented here are to be obtained by machine computation. In all solutions of this type a compromise must be made between the desire for reasonable accuracy and continuity on the one hand and reasonable computation time and cost on the other. Such a compromise is always a matter of engineering judgment and economics. Because of the nature of the problem at hand and the capability of the IBM 1620, the computation time and corresponding cost impose a restriction upon the allowable accuracy and continuity of the present results. However, the main concern in this thesis is with the demonstration of the applicability of the of the Zadah theory in the computational design of a finite memory filter and with the demonstration of the general trend (downward) of the mean square error with increasing memory time. Because these demonstrations can be conducted without taxing the capability of the machine type computer, and because the interpolation and extrapolation of the present results are simply a routine matter, the full power of the design method will not be demonstrated here.

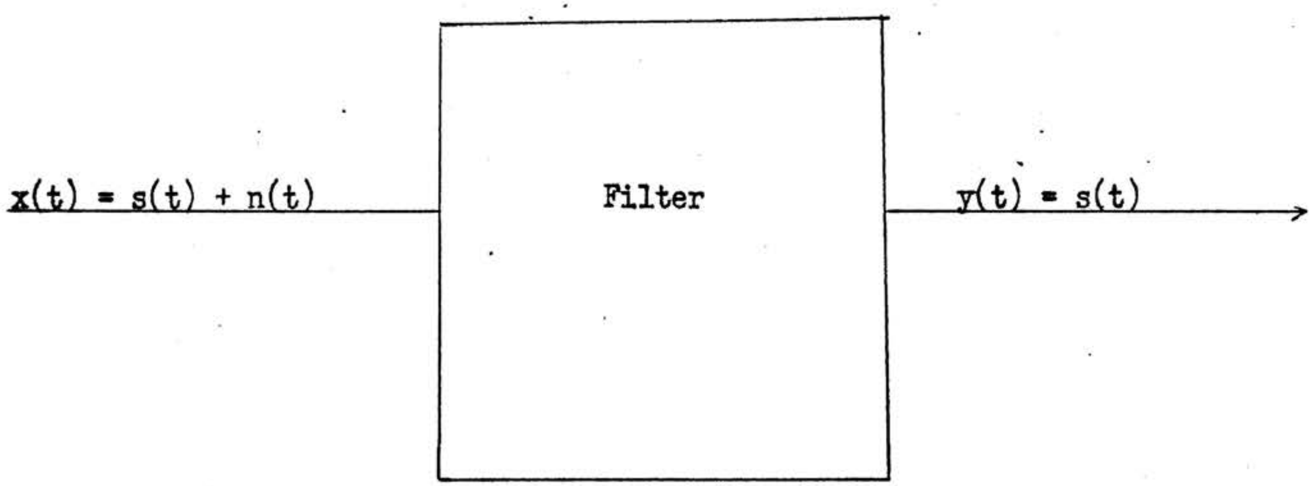

Fig. 1 The System under Consideration

II. THEORETICAL CONSIDERATION

A. Pertinent Terminology

The following definition of symbols and terms will be used throughout the ensuing discussion.

(1) The symbol N is used to denote a nonlinear two pole. By two-pole is meant a system which has only one input and one output terminal.

(2) The symbols $x(t)$ and $y(t)$ are used to denote the system input and output respectively.

(3) A two pole N is said to be additive if for any two time functions $x_1(t)$ and $x_2(t)$ in the input space of N, the following relation holds

$$N(X_1 + X_2) = N(X_1) + N(X_2)$$

In other words, the principle of superposition holds.

(4) A two pole N is said to be homogeneous if for any time function $x(t)$ in the input space of N and any real constant α, the following relation holds

$$N(\alpha X) = \alpha N(X)$$

(5) A two pole that is both additive and homogeneous is then said to be linear.

(6) A non linear two pole is one that is either nonadditive or nonhomogeneous or both.

(7) The class η_1, η_2, η_3---- are defined as the collections of all filters whose input-output relationships can be expressed in the following forms:

$$\eta_1;\ y(t) = \int_0^T K\left[x(t-\tau_1),\tau_1\right] d\tau_1$$

$$\eta_2;\ y(t) = \int_0^T\int_0^T K\left[x(t-\tau_1),\ x(t-\tau_2),\tau_1,\tau_2\right] d\tau_1\ d\tau_2$$

$$\eta_3;\ y(t) = \int_0^T\int_0^T\int_0^T K\left[x(t-\tau_1),\ x(t-\tau_2),\ x(t-\tau_3),\tau_1,\tau_2,\tau_3\right] d\tau_1 d\tau_2 d\tau_3$$

. . .

B. Optimization

We confine our attention to the invariant finite state filter. Let the input x(t) to the filter consist of Poisson distributed signal s(t) plus Guassian distributed noise n(t). The filter is finite memory with time duration T. The output can be defined as

$$y(t) = \int_0^T K[x(t-\tau),\tau]\,d\tau$$

According to Zadah's definition, this is of class η_1.

We assume that the s(t) and n(t) are independent and additive stationary random processes. We denote the sum as x(t):

$$x(t) = s(t) + n(t)$$

Denoting the desired output by $y_d(t)$ and indicating the characteristic function of the filter which yields it by $K_d[s(t-\tau),\tau]$ we have

$$y_d(t) = \int_{-\infty}^{\infty} K_d[s(t-\tau),\tau]\,d\tau$$

If the desired output is $y_d(t) = s(t)$ then

$$K_d[s(t-\tau),\tau] = s(t-\tau)\cdot\delta(\tau)$$

where $\delta(\tau)$ is a Dirac Delta Function. By definition the error is

$$e(t) = y(t) - y_d(t)$$

For convenience, the ensemble average of e(t) is assumed to be zero, or $\overline{e(t)} = 0$. The quantity of interest is the average of $e^2(t)$ or $\overline{e^2(t)}$, and the optimum filter within class η_1 is that filter which minimizes the $\overline{e^2(t)}$.

$$\overline{e^2(t)} = \overline{[y(t) - y_d(t)]^2}$$

$$= \overline{\left\{ \int_0^T K[x(t-\tau),\tau]\,d\tau - \int_{-\infty}^{\infty} K_d[s(t-\tau),\tau]\,d\tau \right\}^2}$$

$$= \overline{\int_0^T\int_0^T K[x(t-\tau_1),\tau_1] \cdot K[x(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2}$$

$$-2\overline{\int_{-\infty}^{\infty}\int_0^T K[x(t-\tau_1),\tau_1] \cdot K_d[s(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2}$$

$$+\overline{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d[s(t-\tau_1),\tau_1] \cdot K_d[s(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2} \qquad (1)$$

The last term of Eq. (1) does not involve $K[x(t-\tau),\tau]$ and therefore is irrevelant to the determination of the minimization of the characteristic function $K[x(t-\tau),\tau]$. The quantity that needs to be minimized therefore consists of the first two terms in Eq. (1) and this quantity will be denoted by F.

The following notation is used for convenience:

$x_1 = x(t-\tau_1)$

$x_2 = x(t-\tau_2)$

$s_1 = s(t-\tau_1)$

$s_2 = s(t-\tau_2)$

$P(x_1, x_2;\tau_1-\tau_2)$ = The joint probability density function between x_1 and x_2

The quantity F to be minimized is

$$F = \overline{\int_0^T\int_0^T K[x_1,\tau_1] \cdot K[x_2,\tau_2]\,d\tau_1 d\tau_2}$$

$$-2\overline{\int_{-\infty}^{\infty}\int_0^T K[x_1,\tau_1] \cdot K_d[s_2,\tau_2]\,d\tau_1 d\tau_2}$$

Expressing the average as an ensemble average, in terms of probability density functions we have

$$F = \int_0^T\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty} K[x_1, \tau_1]\,K[x_2, \tau_2]\,P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_1 d\tau_1 dx_2 d\tau_2$$
$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty} K[x_1, \tau_1]\,K_d[s_2, \tau_2]\,P(x_1,\ s_2; \tau_1 - \tau_2)\,dx_1 d\tau_1 ds_2 d\tau_2 \tag{2}$$

By calculating the variation in F due to an admissible variation $\epsilon\Delta K$ in $K[x(t-\tau), \tau]$ we can minimize the quantity F. In Eq. (2) replace $K[x(t-\tau), \tau]$ by $K + \epsilon\Delta K$ and let F take the form

$$F = \int_0^T\int_{-\infty}^{\infty} K_1\,dx_1 d\tau_1 \int_0^T\int_{-\infty}^{\infty} K_2 P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$
$$-2\int_0^T\int_{-\infty}^{\infty} K_1\,dx_1 d\tau_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2 P(x_1,\ s_2; \tau_1 - \tau_2)\,ds_2 d\tau_2$$
$$= \int_0^T\int_{-\infty}^{\infty} K_1\Big[\int_0^T\int_{-\infty}^{\infty} K_2 P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$
$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d P(x_1,\ s_2; \tau_1 - \tau_2)\,ds_2 d\tau_2\Big]\,dx_1 d\tau_1$$

Where K_i represents $K[x(t-\tau_i), \tau_i]$; $i = 1, 2$. Thus,

$$F + \Delta F$$
$$= \int_0^T\int_{-\infty}^{\infty} (K_1 + \epsilon\Delta K_1)\Big[\int_0^T\int_{-\infty}^{\infty} (K_2 + \epsilon\Delta K_2) P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$
$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d P(x_1,\ s_2; \tau_1 - \tau_2)\,ds_2 d\tau_2\Big]\,dx_1 d\tau_1 \tag{3}$$

By subtracting Eq. (2) from Eq. (3) we obtain

$$\Delta F = \int_0^T\int_{-\infty}^{\infty} \epsilon\Delta K_1\Big[\int_0^T\int_{-\infty}^{\infty} (K_2 + \epsilon\Delta K_2) P(x_1, x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$
$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d P(x_1,\ s_2; \tau_1 - \tau_2)\,ds_2 d\tau_2\Big]\,dx_1 d\tau_1$$
$$+\int_0^T\int_{-\infty}^{\infty} K_1\,dx_1 d\tau_1 \int_0^T\int_{-\infty}^{\infty} \epsilon\Delta K_2 P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$
$$= \epsilon\int_0^T\int_{-\infty}^{\infty} \Delta K_1\,dx_1 d\tau_1 \int_0^T\int_{-\infty}^{\infty} K_2 P(x_1,\ x_2; \tau_1 - \tau_2)\,dx_2 d\tau_2$$

$$=\epsilon\int_0^T\int_{-\infty}^{\infty}K_1 dx_1 d\tau_1\int_0^T\int_{-\infty}^{\infty}\Delta K_2 P(x_1,\ x_2;\tau_1-\tau_2)dx_2 d\tau_2$$

$$-2\epsilon\int_0^T\int_{-\infty}^{\infty}\Delta K_1 dx_2 d\tau_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_d P(x_1,s_2;\tau_1-\tau_2)ds_2 d\tau_2$$

$$+\epsilon^2\int_0^T\int_{-\infty}^{\infty}\Delta K_1 dx_1 d\tau_1\int_0^T\int_{-\infty}^{\infty}\Delta K_2 P(x_1,\ x_2;\tau_1-\tau_2)dx_2 d\tau_2$$

$$=2\epsilon\int_0^T\int_{-\infty}^{\infty}\Delta K_1\Big[\int_0^T\int_{-\infty}^{\infty}K_2 P(x_1,\ x_2;\tau_1-\tau_2)dx_2 d\tau_2$$

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_d P(x_1,\ s_2;\tau_1-\tau_2)dx_2 d\tau_2\Big]dx_1 d\tau_1$$

$$+\epsilon^2\int_0^T\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty}\Delta K_1\Delta K_2 P(x_1,\ x_2;\tau_1-\tau_2)dx_1 d\tau_1 dx_2 d\tau_2$$

Setting $\left.\dfrac{\partial(\Delta F)}{\partial\epsilon}\right|_{\epsilon=0}=0$ yields

$$\int_0^T\int_{-\infty}^{\infty}K[x_1,\tau_1]\Big\{\int_0^T\int_{-\infty}^{\infty}K[x_2,\tau_2]P(x_1,\ x_2;\tau_1-\tau_2)dx_2 d\tau_2$$

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_d[s_2,\tau_2]P(x_1,\ s_2;\tau_1-\tau_2)ds_2 d\tau_2\Big\}dx_1 d\tau_1=0 \qquad (4)$$

From Eq. (3) it can be shown that $\dfrac{\partial^2(\Delta F)}{\partial\epsilon^2}$ is always positive, so that the quantity F is a minimum when the condition of Eq. (4) is applied. From Eq. (4) we obtain

$$\int_0^T\int_{-\infty}^{\infty}K[x_2,\tau_2]P(x_1,\ x_2;\tau_1-\tau_2)dx_2 d\tau_2$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_d[s_2,\tau_2]P(x_1,\ s_2;\tau_1-\tau_2)ds_2 d\tau_2 \qquad (5)$$

where $\tau_1\geq 0$

Eq. (5) is the Wiener-Hopf type equation for the class η_1; the characteristic function of the optimum filter of class η_1 is the solution of this linear integral equation.

Now recall that our desired output is assumed to be s(t), the signal itself. Thus

$$\int_{-\infty}^{\infty} K_d[s(t-\tau),\tau]\,d\tau = s(t) \tag{6}$$

or $K_d[s, t] = s \cdot \delta(t)$

By substituting Eq. (6) into Eq. (5)

$$\int_0^T\int_{-\infty}^{\infty} K[x_2,\tau_2]P(x_1,\ x_2;\tau_1-\tau_2)\,dx_2\,d\tau_2$$

$$=\int_{-\infty}^{\infty} sP(x_1,\ s;\tau_1)\,ds \tag{7}$$

If we now assume s(t) and n(t) are discrete-amplitude processes, the joint probability density function $P(x_1,\ x_2;\tau_1-\tau_2)$ and $P(x_1,\ s;\tau_1)$ can be expressed in terms of delta functions.

III. APPLICATION

A. The Finite Memory Solution

The signal and noise situation which we consider here was actually considered first by Zadah (8). However, his treatment was confined to a theoretical development and he did not actually solve the filter problem per se. Instead he left the result in a form which could be solved only by machine type computation.

To begin our solution for $K[x_2, \tau_2]$ we will now specify the signal and noise statistics. The noise is assumed to be Guassian with correlation function $\phi_n(\tau) = e^{-\alpha|\tau|}$ where α is a positive constant. The second probability distribution function is given by

$$P_n(n_1, n_2) = \frac{1}{2\left[1-\exp(-2\alpha|\tau_1-\tau_2|)\right]^{\frac{1}{2}}}$$
$$\cdot \exp\left\{-\frac{n_1^2 - 2n_1n_2\exp(-\alpha|\tau_1-\tau_2|) + n_2^2}{2\left[1 - \exp(-2\alpha|\tau_1-\tau_2|)\right]}\right\} \qquad (8)$$

where $n_1 = n(t-\tau_1)$, $n_2 = n(t-\tau_2)$. The signal s(t) is taken to be a random square wave which takes on the value 1 or 0 with equal probability over intervals of random length D, the probability density of D being $\beta e^{-\beta D}$ (Poisson Process) where $1/\beta = \langle D \rangle_{av}$. It can readily be shown* that the probability that

---

* R. L. Sternberg, Bulletin American Mathematics Society. 58, 389 (1952), abstract only.

$t-\tau_1$ and $t-\tau_2$ lie within the same interval is

$$Q(\tau_1-\tau_2) = \exp(-\beta|\tau_1-\tau_2|)$$

The second probability distribution function is given (see Zadah, op. cit.) by

$$P_s(s_1, s_2) = \frac{1}{4}\left[1-e^{-\beta|\tau_1-\tau_2|}\right]\left[\delta(s_2-1)\delta(s_1) + \delta(s_1-1)\delta(s_2)\right]$$
$$+ \frac{1}{4}\left[1+e^{-\beta|\tau_1-\tau_2|}\right]\left[\delta(s_1)\delta(s_2) + \delta(s_1-1)\delta(s_2-1)\right] \tag{9}$$

where $s_1 = s(t-\tau_1)$, $s_2 = s(t-\tau_2)$,

and where $\delta(s)$ is the delta function.

In order to solve the characteristic function of the finite memory nonlinear filter we let the input $x(t)$ be equal to $s(t) + n(t)$ and the desired output $y_d(t)$ equal to $s_2(t)$. Then $K_d[s_2, \tau_2]$ can be written as $s_2\delta(\tau_2)$. By substituting Eq. (5) and by denoting the right hand side of Eq. (5) by the symbol R, we obtain

$$R = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d[s_2, \tau_2]P(x_1, s_2; \tau_1-\tau_2)ds_2 d\tau_2$$
$$= \int_{-\infty}^{\infty} s\, P(x_1, s; \tau_1)ds \tag{7}$$

However,

$$P(x_1, s; \tau_1) = \int_{-\infty}^{\infty} P_s(s_1, s)P_n(x_1 - s_1)ds_1 \tag{10}$$

Also for $s_2 = s$, Eq. (9) yields

$$P_s(s_1, s) = \frac{1}{4}[1-e^{-\beta|\tau_1|}][\delta(s_1)\delta(s-1) + \delta(s_1-1)\delta(s)]$$

$$+ \frac{1}{4}[1+e^{-\beta|\tau_1|}][\delta(s_1)\delta(s) + \delta(s_1-1)\delta(s-1)] \tag{11}$$

$$P_n(x_1 - s_1) = 1/(2\pi)^{\frac{1}{2}} \exp[-(x_1 - s_1)^2/2] \tag{12}$$

Substituting Eqs. (10), (11), and (12) into Eq. (7), we obtain:

$$R = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s\Big\{\frac{1}{4}[1-e^{-\beta|\tau_1|}][\delta(s_1)\delta(s-1) + \delta(s_1-1)\delta(s)]$$

$$+ \frac{1}{4}[1+e^{-\beta|\tau_1|}][\delta(s_1)\delta(s) + \delta(s_1-1)\delta(s-1)]\Big\}$$

$$\cdot\ 1/(2\pi)^{\frac{1}{2}} \exp\Big[-\frac{(x_1-s_1)^2}{2}\Big] ds_1 ds$$

or

$$R = 1/4(2\pi)^{\frac{1}{2}}\Big\{(1-e^{-\beta|\tau_1|})\exp\Big[-\frac{x_1^2}{2}\Big] + (1+e^{-\beta|\tau_1|})$$

$$\cdot\ \exp[-(x_1-1)^2]\Big\} \tag{13}$$

By representing the second order probability distribution functions for the signal and noise as a two-dimensional convolution integral, we have

$$P(x_1, x_2; \tau_1-\tau_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_n(x_1 - s_1, x_2-s_2)P_s(s_1, s_2)ds_1 ds_2 \tag{14}$$

By substituting Eqs. (5) and (9) into Eq. (13), we find

$$P(x_1,\ x_2; \tau_1-\tau_2) = \frac{1-\exp(-\beta|\tau_1-\tau_2|)}{8\pi\left[1-\exp(-2\alpha|\tau_1-\tau_2|)\right]^{\frac{1}{2}}}$$

$$\left\{\exp\left[-\frac{x_1{}^2-2x_1(x_2-1)e^{-\alpha|\tau_1-\tau_2|}+(x_2-1)^2}{2(1-e^{-2\alpha|\tau_1-\tau_2|})}\right]\right.$$

$$+\left.\exp\left[-\frac{(x_1-1)^2-2(x_1-1)x_2e^{-\alpha|\tau_1-\tau_2|}+x_2{}^2}{2(1-e^{-2\alpha|\tau_1-\tau_2|})}\right]\right\}$$

$$+\frac{1+\exp(-\beta|\tau_1-\tau_2|)}{8\pi\left[1-\exp(-2\alpha|\tau_1-\tau_2|)\right]^{\frac{1}{2}}}$$

$$\cdot\left\{\exp\left[-\frac{x_1{}^2-2x_1x_2e^{-\alpha|\tau_1-\tau_2|}+x_2{}^2}{2(1-e^{-2\alpha|\tau_1-\tau_2|})}\right]\right.$$

$$\left.+\exp\left[-\frac{(x_1-1)^2-2(x_1-1)(x_2-1)e^{-\alpha|\tau_1-\tau_2|}+(x_2-1)^2}{2(1-2^{-2\alpha|\tau_1-\tau_2|})}\right]\right\} \qquad (13)$$

By using Eqs. (11) and (13) in Eq. (1) we can solve for the kernel function $K[x,\tau]$. For discrete data and machine computation, this solution reduces to the solution of a set of simultaneous linear algebraic equations.

$$\sum_{i=0}^{I} \sum_{\ell=-L}^{L} K[x_{2_\ell}, \tau_{2_i}] P(x_1, x_2; \tau_1 - \tau_{2_i}) = \int_{-\infty}^{\infty} s \cdot P(x_1, s; \tau_1) ds = f(x_1)$$

It may be noticed that the equations are functions of x_1 only.

In order to solve for the unknown kernel we need I(2L + 1) simultaneous equations. The more equations we use, the more accurate will be our results. Presently, only the IBM 1620 is available. This computer is rather slow in its handling of this type of problem and we are essentially restricted to 64 equations and 64 unknowns. The results of the computer solution are shown in tables 1 through 5.

τ_2 \ x_2	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
1	-2.0845	1474.36	-152.05	-289.99	222.35	-161.12	45.6608	2.54827	-5.42063
1.5	1016168	55335.8	-58904.9	-20489.7	-4006.8	5867.7	-4507.5	4487.8	691.38
2.0	-716692	-174527	-322119	-17123.5	66417.4	308979	-55918	-17813	2990.4
2.5	4326586	-1644496	-316922	134498	12312.9	12398.3	3282.97	8909.53	-20340.4
3.0	386589	208056	128064	222320	-428260	-5778.9	104272	-1928.02	-15391.5
3.5	-205548	55445.4	-54587	540155	10060.6	-32241.6	-78283.8	-2313.9	5066.3
4.0	851592	0747384	-1573.56	-44469.3	-74412.9	776232	-10878.8	2518.1	-9.87405

$\beta = 10$ $\alpha = 1$ $\tau_1 = 0$

Table 1 Characteristic Function $K[x_2, \tau_2]$

τ_2 / x_2	1	3	5	7	9	11	13
-0.4	-4.31325	163769	572178	-377811	844056	119905	-368119
-0.2	-6.67610	-66698.7	-246427	-631882	116715	72308.6	181563
0	1542.51	-215534	-386535	-105245	368820	121642	370903
0.2	-1413.43	45581.9	49626.6	735434	237120	-963161	10688
0.4	1503.71	112264	-128845	194306	404427	386605	4.68567

$\beta = 10 \qquad \alpha = 1 \qquad \tau_1 = 0$

Table 2 Characteristic Function $K[x_2, \tau_2]$

τ_2 \ x_2	-0.4	-0.2	0	0.2	0.4
1	-2.49228	-67.2170	477.806	0280.816	233.446
3	-16277.8	66589.9	6472.11	-94080.5	21081.6
5	-241234	47669.6	197015	67158	153794
7	2311.70	-352213	18425.6	-144142	-51428.7
9	-157556	-69433.2	-120693	-13273.6	-409764
11	-21485.3	12443.5	-43804.4	17021.4	5.00000
13	-100346	-4971.70	257924	31336	6893.03
15	15098.2	-590701	-1496.5	970466	-352.89
17	1194027	29315.3	-631947	-20315.2	3.81694

$\beta = 10 \qquad \alpha = 1 \qquad \tau_1 = 0$

Table 3 Characteristic Function $K[x_2, \tau_2]$

x_2 \ τ_2	1	4	7	10	13	16	19
-0.4	-2.59053	-29760.5	-102738	133242	-106462	-43691.7	-680751.1
-0.2	38311.23	126196	-51506.2	128498.7	220822	-396471	254329
0	-3728.08	36265.2	859320	115639	-50483.1	363804	201030
0.2	2532.16	-297444	-507894	279244	337433	1527921	248487
0.4	-2283.97	948340	240098	-755914	-551328	-37925.7	3.78585

$\beta = 10 \qquad \alpha = 1 \qquad \tau_1 = 0$

Table 4 Characteristic Function $K[x_2, \tau_2]$

$\alpha = 1$ $\beta = 10$ $\tau_1 = 0$ $x = 1$

τ_2	$K[x, \tau_2]$
0.5	.16949014E+01
1.0	.30915959E+02
1.5	.68522904E+03
2.0	.80072826E+04
2.5	.46811498E+05
3.0	.12091136E+06
3.5	.71034544E+05
4.0	.12839460E+06
4.5	.10427309E+05
5.0	.10237671E+06
5.5	.21583012E+06
6.0	.24166547E+05
6.5	.21034894E+04
7.0	.50292116E+05
7.5	.81749133E+05
8.0	.24931044E+06
8.5	.34685047E+06
9.0	.82873935E+05
9.5	.30520050E+04
10.0	.25772759E+06
10.5	.15531366E+05
11.0	.25421551E+06
11.5	.60767196E+05
12.0	.32812884E+06
12.5	.43492501E+06
13.0	.37047647E+06
13.5	.40004805E+05
14.0	.14642160E+06
14.5	.65359103E+06
15.0	.18384386E+06
15.5	.85515320E+05
16.0	.71019181E+06
16.5	.80706402E+05

Table 5 Characteristic $K[x_2, \tau_2]$ Varies with τ Alone

B. The Mean Square Error

The mean square error is expressed as

$$\overline{e^2(t)} = \overline{[y_d(t)-y(t)]^2} = \overline{y_d{}^2(t)-2y_d(t)\cdot y(t) + y^2(t)}$$

where

$$y_d(t) = \int_{-\infty}^{\infty} K_d[s(t-\tau),\tau]\, d\tau$$

and

$$y(t) = \int_0^T K[x(t-\tau),\tau]\, d\tau$$

Thus

$$\overline{e^2(t)} = \overline{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d[s(t-\tau_1),\tau_1]K_d[s(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2}$$
$$-2\overline{\int_{-\infty}^{\infty}\int_0^T K[x(t-\tau_1),\tau_1]K_d[s(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2}$$
$$+\overline{\int_0^T\int_0^T K[x(t-\tau_1),\tau_1]K[x(t-\tau_2),\tau_2]\,d\tau_1 d\tau_2}$$

Since we assumed the desired output is s(t) itself, the first term is exactly equal to $\overline{s^2}(t)$. For convenience, let the sum of the second term and the third term be called I. Then,

$$I = \int_0^T\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty} K[x_1,\tau_1]\,K[x_2,\tau_2]P(x_1,\ x_2;\tau_1-\tau_2)dx_1 d\tau_1 dx_2 d\tau_2$$

$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty} K[x_1,\tau_1]K_d[s_2,\tau_2]P(x_1,\ s_2;\tau_1-\tau_2)dx_1 d\tau_1 ds_2 d\tau_2$$

$$= \int_0^T\int_{-\infty}^{\infty} K[x_1, \tau_1]\Big\{\int_0^T\int_{-\infty}^{\infty} K[x_2, \tau_2]P(x_1,\ x_2; \tau_1 - \tau_2)dx_2 d\tau_2$$

$$-2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d[s_2, \tau_2]P(x_1,\ s_2; \tau_1 - \tau_2)ds_2 d\tau_2\Big\}dx_1 d\tau_1$$

However, we have shown that when

$$\int_0^T\int_{-\infty}^{\infty} K[x_2, \tau_2]\, P(x_1,\ x_2; \tau_1 - \tau_2)dx_2 d\tau_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_d[s_2, \tau_2]P(x_1,\ s_2; \tau_1 - \tau_2)ds_2 d\tau_2$$

the filter yields its minimum mean square error.. Thus when we substitute this information into I we obtain

$$I = \int_0^T\int_{-\infty}^{\infty} K[x_1, \tau_1]\Big[-\int_0^T\int_{-\infty}^{\infty} K[x_2, \tau_2]P(x_1,\ x_2; \tau_1 - \tau_2)dx_2 d\tau_2\Big]dx_1 d\tau_1$$

Therefore,

$$\overline{e^2(t)} = \overline{s^2(t)} - \int_0^T\int_{-\infty}^{\infty}\int_0^T\int_{-\infty}^{\infty} K[x_1, \tau_1]K[x_2, \tau_2]P(x_1,\ x_2; \tau_1 - \tau_2)dx_1 d\tau_1 dx_2 d\tau_2$$

$$= \overline{s^2(t)} - \overline{\Big[\int_0^T K[x, \tau]d\tau\Big]^2} \qquad (14)$$

The signal power $\overline{s^2(t)}$ is always a constant value, so the mean square error $\overline{e^2(t)}$ is essentially just the function $\overline{\left[\int_0^T K[x, \tau]d\tau\right]^2}$.

The memory time T of a filter is defined as the duration of that portion of the "past" of the input which the filter uses to determine the "present" value of the filter output. The greater the memory time, T, the greater will be the amount of the input's "past" which is used to determine the "present" output. In other words, as T increases, the filter can be looked upon as doing more and more of a "post mortem" on the input to yield the "present" output. Intuitively, one would expect that as more and more of the "past" is considered, the better might be the filter's ability to separate signal from noise. At least one would expect that the filter would do at least as good a job for memory time T_2 as for T_1 if $T_2 > T_1$. In other words, one would intuitively expect that the mean square error $\overline{e^2(t, T)}$ would be a monotonic decreasing function of T. However, no general proof of this property has been found in the literature, and so it is interesting to investigate the behavior of $\overline{e^2(t, T)}$ vs. T for the filter under consideration here. Such a study was performed with the aid of the IBM 1620 and the results are plotted in Fig. 3 to 6. Although it might be expected that the downward trend of the curve should become asymptotic to some minimum value, the limitations of the IBM 1620 precluded the determination of such a value.

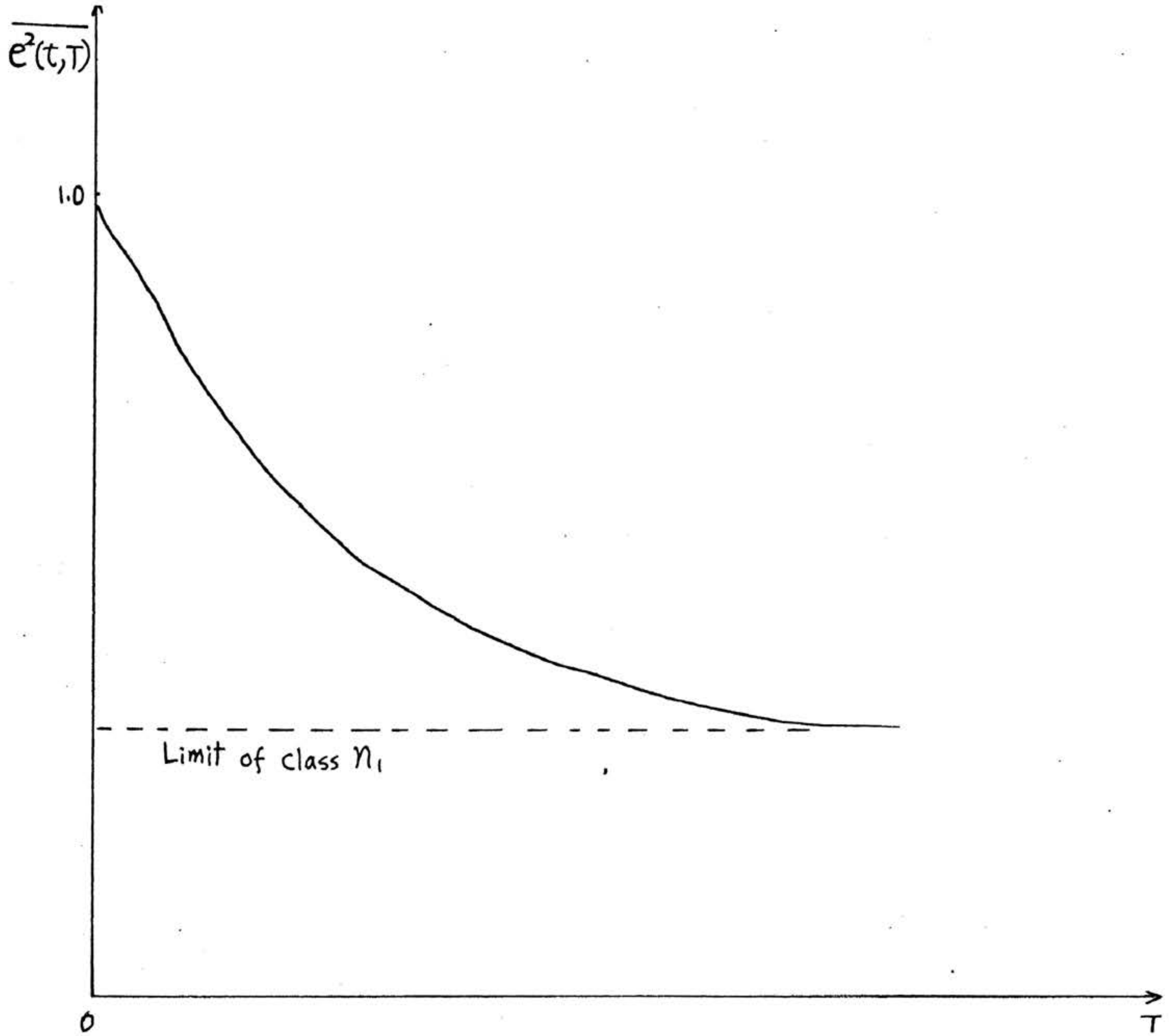

Fig. 2 Mean Square Error by Theoretical Consideration

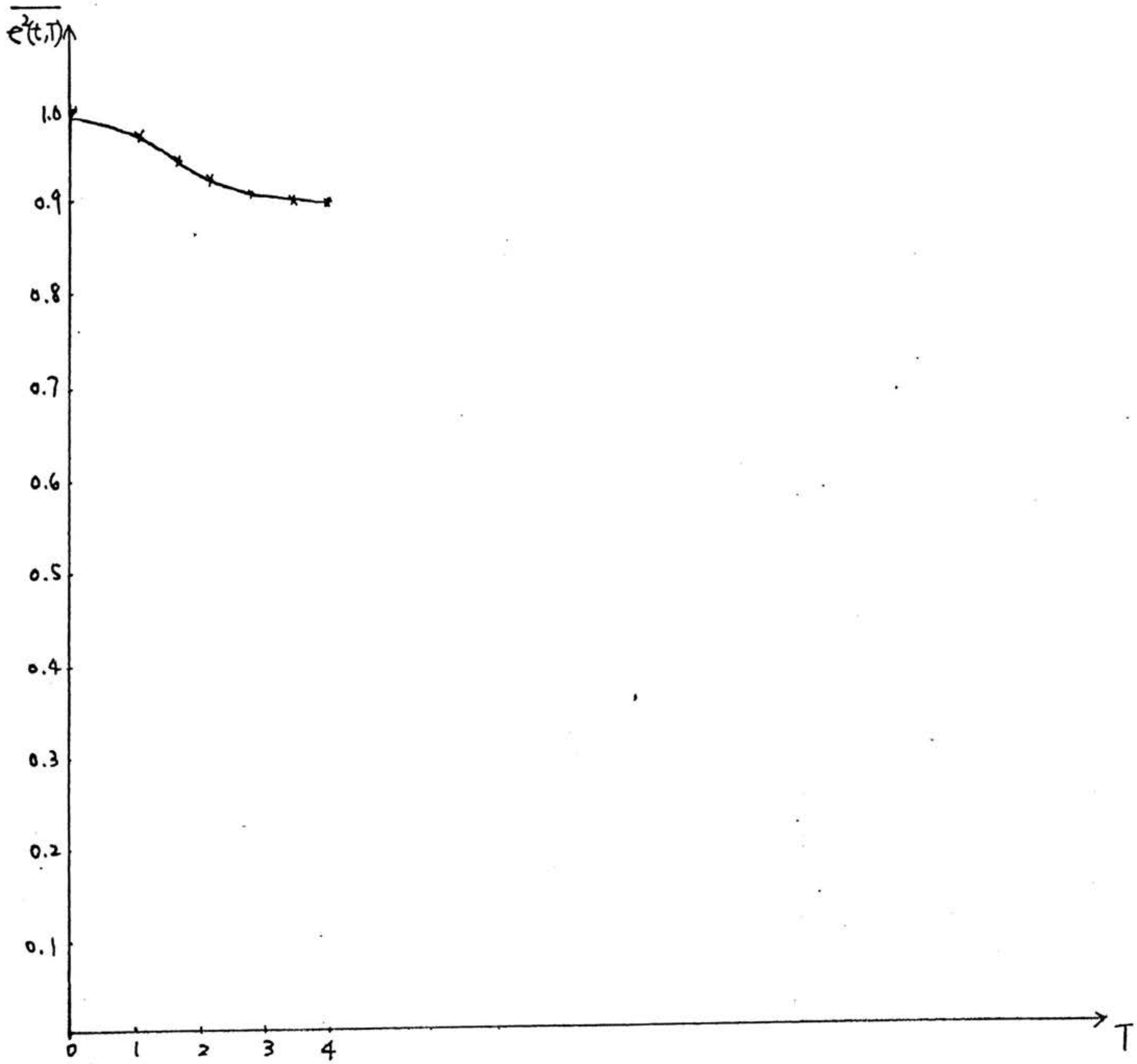

Fig. 3 Actual Computation of Mean Square Error from Table 1

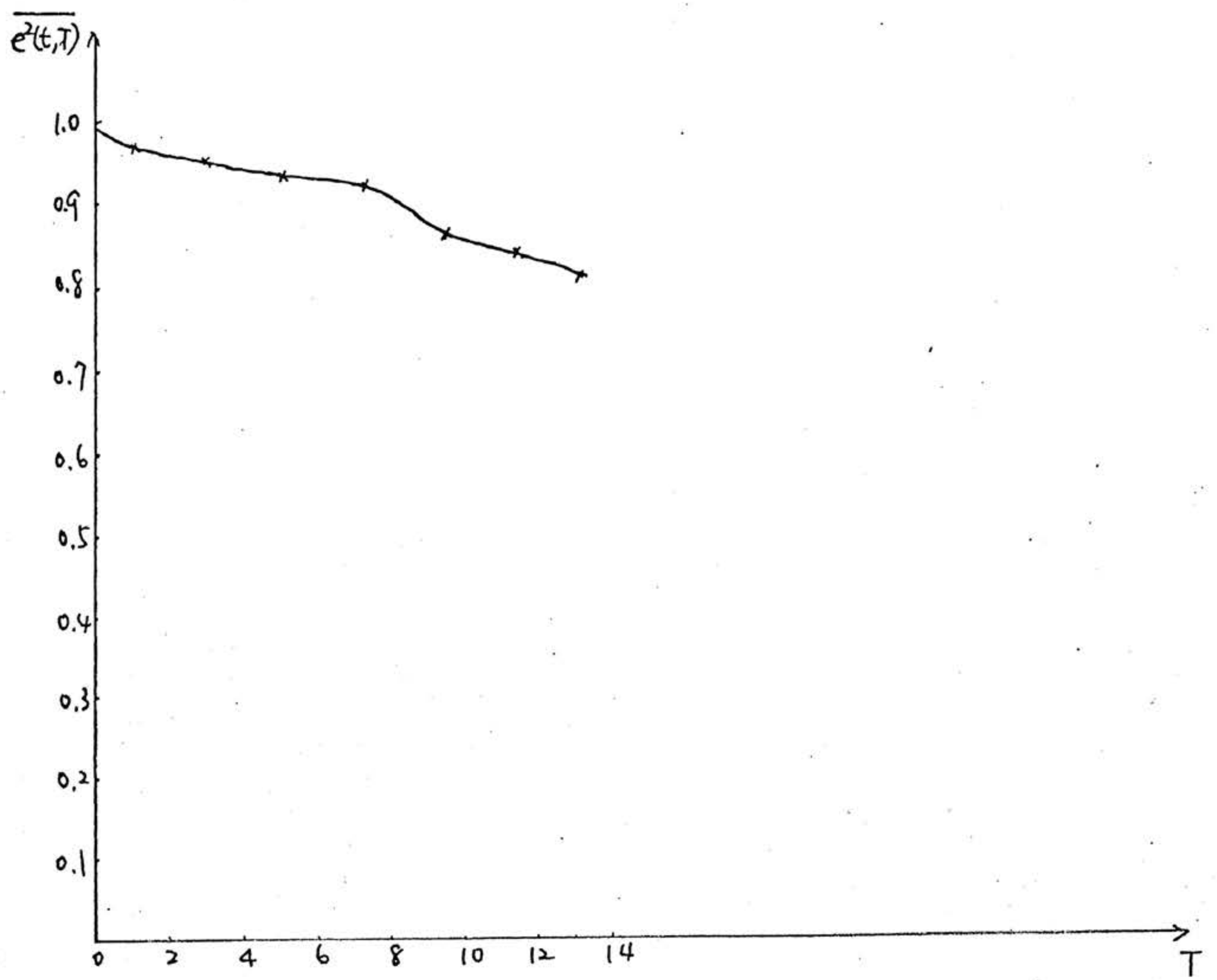

Fig. 4 Actual Computation of Mean Square Error from Table 2

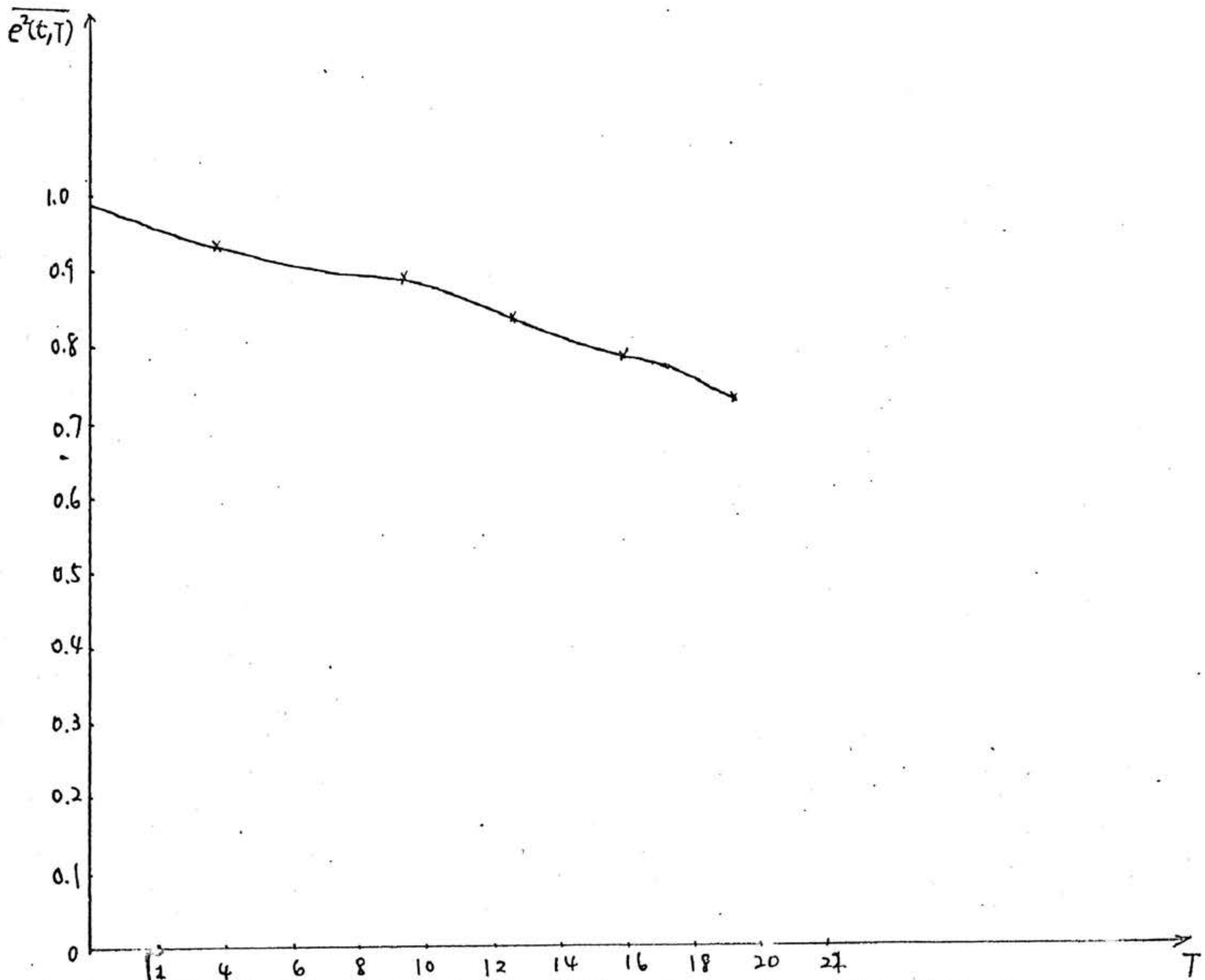

Fig. 5 Actual Computation of Mean Square Error from Table 3

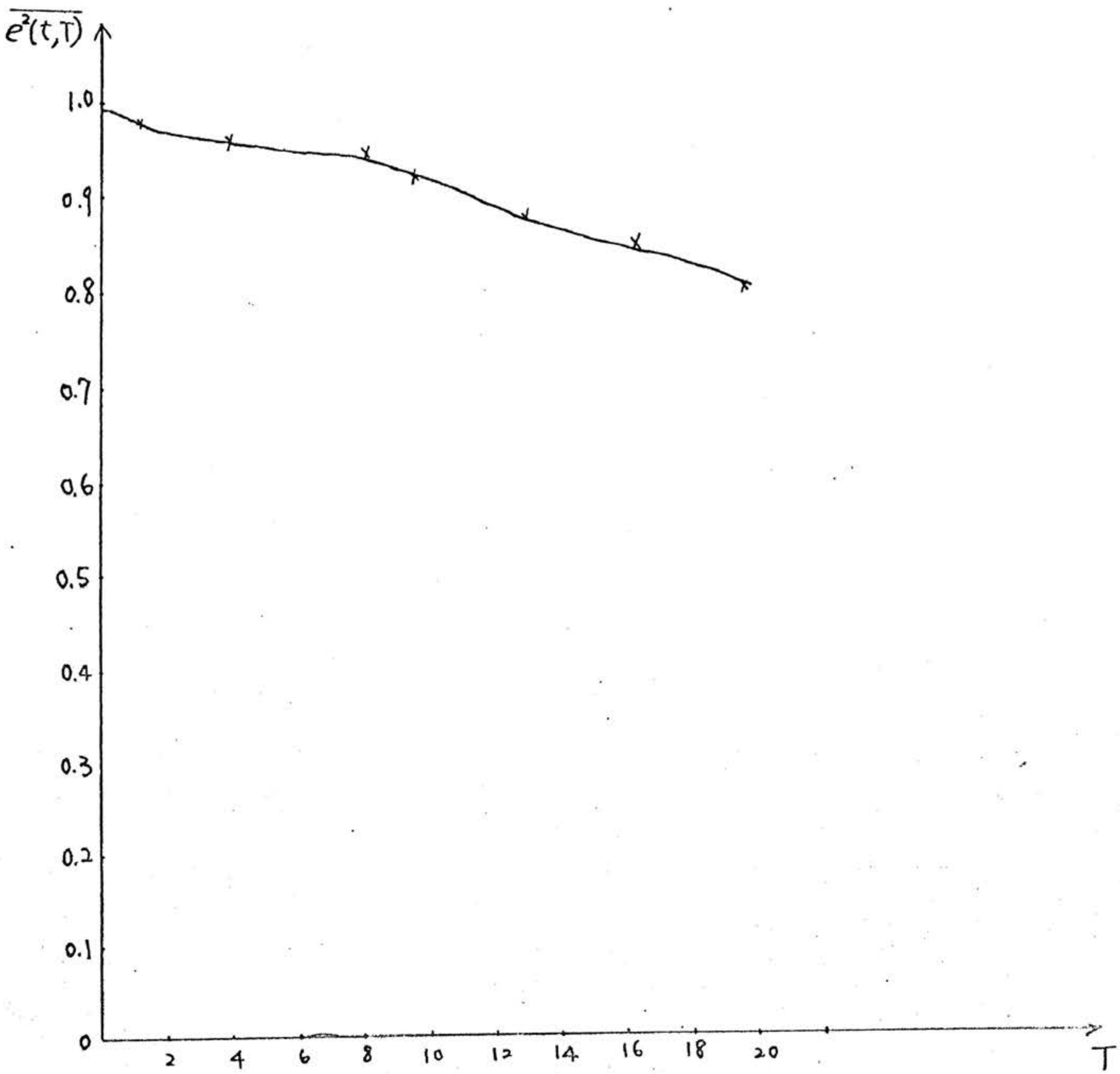

Fig. 6 Actual Computation of Mean Square Error from Table 4

IV. CONCLUSIONS

On the basis of the study performed here, it is clear that it is possible to use the Zadah theory and a machine type computer to design an optimal finite memory nonlinear filter for the extraction of Poisson distributed signal pulses from a background of Gaussian noise. The result of the design is a two-dimensional kernel function which can be used to operate upon the input data (signal plus noise) to produce a minimum mean square estimate of the signal. Examples of several such operators are shown in Tables 1 through 4. Although the resultory kernel functions appear rather oscillatory, it is always possible (at least in principle) to obtain a smoother and more continuous variation by performing calculations at a large number of points in-between those which were taken in this thesis. A demonstration of this property is shown by the results which are tabulated in Table 5.

Because this thesis is concerned more witha demonstration of the validity of the design method than it is with the actual processing of data, and because the IBM 1620 is greatly limited in its ability to solve for these two-dimensional kernels, the full power of the design method in terms of its ability to yield continuous results is not exploited here. For this reason the kernel values determined here will not actually be used with input data to extract the original signal. There simply is not enough fine-grain character to the results obtained to make such an estimation interesting. Nevertheless, it is obvious from the work presented here that such a procedure is possible and can be obtained in a routine manner in any situation which justifies the computer time and the corresponding cost.

It is also clear on the basis of the present study that the intuitive value of a monotonic decreasing variation of minimum mean square error with increasing memory time is correct. This behavior is illustrated in Fig. 3 to 6. Here again a finer-grain analysis would be even more convincing than the one shown here. However, the cost of such an analysis on the IBM 1620 is deemed prohibitive, particularly since our present concern has been more with a demonstration of the trend than with a highly detailed investigation of the actual behavior of $\overline{e^2(t, T)}$ vs. T.

REFERENCES

(1) Wiener, N. "Extrapolation, Interpolation and Smoothing Time Series" Wiley, 1949.

(2) Wiener, N. "Response of a Nonlinear Device to Noise" M.I.T. Radiation Laboratory Report No. 129, 1942.

(3) Wiener, N. "Nonlinear Problem in Random Theory" Wiley, 1958.

(4) Lee, Y. W. "Statistical Theory of Communication" John Wiley & Sons Inc. 1960.

(5) Bose, A. G. "A Theory of Nonlinear System" M.I.T. Research Laboratory of Electronic Report No. 309, 1956.

(6) Singleton, H. E. "Theory of Nonlinear Transducers" M.I.T. Technical Report No. 309.

(7) Zadah, L. A. "A Contribution to the Theory of Nonlinear Systems" Journal of Franklin Institute, 1953, 255, pp. 387.

(8) Zadah, L. A. "Optimum Nonlinear Filters" Journal of Applied Physics, 1953, 24, pp. 396.

(9) Wang, Benjamin Shaw-hu. "Optimum Zero-Memory Nonlinear" Library of University of Missouri at Rolla, 1964, T 1565.

(10) Tung, F. and Schwarz, R. J. "Optimum Nonlinear Filter for Quantized Inputs" IRE Transaction on Information Theory, Vol. IT-7, pp. 257.

(11) Gabor, F. R. S., Wilby, W. P. L., and Woodcock, R. "A Universal Nonlinear Filter, Predictor and Simulator which Optimizes Itself by a Learning Process" IEE Proceedings, Vol. 108, 1961, pp. 422.

(12) Zadah, L. A. "An Extension of Wiener's Theory of Prediction" Journal of Applied Physics, 1950, 21, pp. 645.

VITA

Ching-Chyoun Yang was born on January 22, 1940, at Taichung, Taiwan, China. He received his primary education there. Then he attended Taiwan Provincial Taichung First Middle School, Taichung, Taiwan, from 1952 to 1958. He then enrolled at the National Taiwan University in September, 1958. He received the Bachelor of Science degree in Electrical Engineering in June, 1962. From July, 1962, to July, 1963, he served as a teaching assistant in the Electrical Engineering Department, Chinese Naval Academy, Tsoying, Taiwan.

In February, 1964, he entered the University of Missouri at Rolla, where he is presently enrolled as a graduate student. He is a member of Kappa Mu Epsilon and IEEE.